Research, Society and Development, v. 11, n. 15, e01111536173, 2022
(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v11i15.36173

Aplicação da modelagem preditiva via árvore de decisão nos casos de Síndrome Respiratória Aguda Grave (SRAG), com ênfase na Corona Virus Disease 2019 (COVID-19) no Brasil referente ao período de 2020 a 2022

Application of predictive modeling via decision tree in cases of Severe Acute Respiratory Syndrome (SARS), with emphasis on Corona Virus Disease 2019 (COVID-19) in Brazil for the period from 2020 to 2022

Aplicación de Modelado Predictivo Vía Árbol de Decisión en casos de Síndrome Respiratorio Agudo Severo (SRAG), con énfasis en la Enfermedad por Corona Virus 2019 (COVID-19) en Brasil para el período de 2020 a 2022

Recebido: 05/10/2022 | Revisado: 17/10/2022 | Aceitado: 18/10/2022 | Publicado: 12/11/2022

Miriam Lecília Farias Ribeiro
ORCID: https://orcid.org/0000-0002-6439-2563
Universidade de Pernambuco, Brasil
E-mail: miriam.ribeiro@upe.br
Natália Moraes Cordeiro
ORCID: https://orcid.org/0000-0001-5294-1353
Universidade Federal Rural de Pernambuco, Brasil
E-mail: nataliamcorddeiro@gmail.com
Dâmocles Aurélio Nascimento da Silva Alves
ORCID: https://orcid.org/0000-0002-7928-1276
Universidade de Pernambuco, Brasil
E-mail: damocles.aurelio@upe.br

Resumo

A Síndrome Respiratória Aguda Grave (SRAG) abrange casos de Síndrome Gripal (SG) que evoluem com comprometimento da função respiratória o que, na maioria dos casos, leva à hospitalização. A pandemia provocada pela Corona Virus Disease (COVID-19) tornou-se o novo desafio mundial. Pacientes que apresentavam determinadas doenças crônicas tinham um prognóstico agravado quando eram apresentados ao novo Coronavírus. É imprescindível determinar os principais grupos de risco para qualquer doença, posto que facilita a tomada de decisão dos profissionais da saúde. Nesta pesquisa objetivou-se aplicar a modelagem preditiva via árvore de decisão (*decision tree*) para estimar a probabilidade do indivíduo que: tenha SRAG ser curado ou ir a óbito e tenha SRAG ser curado ou ir a óbito devido à contaminação e não contaminação por COVID-19, analisando por fim os resultados (casos registrados no Brasil). Essas informações ajudarão os profissionais da saúde a entender como cada comorbidade se comportou/a. Os principais resultados mostraram que o modelo proposto se ajusta bem, encontrando as seguintes porcentagens de sobrevivência: é melhor que o indivíduo que apresentou sintomas da SRAG tenha doença renal e asma do que não tenha comorbidade, pois a chance de cura é maior 7%; é melhor que o indivíduo que apresentou sintomas da SRAG devido à contaminação por COVID-19 tenha doença neurológica, cardiovascular e hematológica do que não tenha comorbidade, pois a chance de cura é maior 14% e por fim, o indivíduo que apresentou sintomas da SRAG, mas que não tenha sido contaminado por COVID-19 tem 75% chance de cura.
Palavras-chave: Árvore de decisão; COVID-19; Estatística; Previsão; SRAG.

Abstract

Severe Acute Respiratory Syndrome (SARS) covers cases of Influenza Syndrome (GS) that evolve with compromised respiratory function which, in most cases, leads to hospitalization. The pandemic caused by the Corona Virus Disease (COVID-19) has become the new global challenge. Patients who had certain chronic diseases had a worse prognosis when they were introduced to the new coronavirus. It is essential to determine the main risk groups for any disease, since it facilitates the decision-making of health professionals. This research aimed to apply predictive modeling via decision tree to estimate the probability of the individual who: has SARS being cured or dying and has SARS being cured or dying due to contamination and not contamination by COVID -19, finally analyzing the results (cases registered in Brazil). This information will help healthcare professionals understand how each comorbidity behaved.

The main results showed that the proposed model fits well, finding the following survival percentages: it is better for the individual who presented symptoms of SARS to have kidney disease and asthma than to have no comorbidity, as the chance of cure is 7% higher; it is better for the individual who presented symptoms of SARS due to contamination by COVID-19 to have neurological, cardiovascular and hematological disease than to have no comorbidity, as the chance of cure is 14% higher and, finally, the individual who presented symptoms of SARS , but who has not been infected by COVID-19 has a 75% chance of cure.
Keywords: Decision tree; COVID-19; Statistic; Prediction; SARS.

Resumen
El Síndrome Respiratorio Agudo Severo (SARS) cubre los casos de Síndrome Influenza (SG) que evolucionan con compromiso de la función respiratoria que, en la mayoría de los casos, conduce a la hospitalización. La pandemia provocada por la Enfermedad del Corona Virus (COVID-19) se ha convertido en el nuevo reto mundial. Los pacientes que tenían ciertas enfermedades crónicas tenían un peor pronóstico cuando se les presentó el nuevo coronavirus. Es fundamental determinar los principales grupos de riesgo de cualquier enfermedad, ya que facilita la toma de decisiones de los profesionales sanitarios. Esta investigación tuvo como objetivo aplicar modelos predictivos vía árbol de decisión para estimar la probabilidad de que el individuo que: tiene SARS se cure o muera y tenga SARS se cure o muera por contaminación y no contaminación por COVID-19, analizando finalmente los resultados (casos registrados en Brasil). Esta información ayudará a los profesionales de la salud a comprender cómo se comportó cada comorbilidad. Los principales resultados mostraron que el modelo propuesto se ajusta bien, encontrando los siguientes porcentajes de supervivencia: es mejor que el individuo que presentó síntomas de SARS tenga enfermedad renal y asma que no tener comorbilidad, ya que la probabilidad de curación es un 7% mayor; es mejor que el individuo que presentó síntomas de SARS por contaminación por COVID-19 tenga enfermedad neurológica, cardiovascular y hematológica que no tener comorbilidad, ya que la probabilidad de curación es un 14% mayor y, finalmente, el individuo que presentó síntomas del SARS, pero quien no ha sido infectado por COVID-19 tiene un 75% de posibilidades de curación.
Palabras clave: Árbol de decisón; COVID-19; Estadística; Predicción; SARS.

1. Introdução

A Síndrome Respiratória Aguda Grave (SRAG) abrange casos de Síndrome Gripal (SG) que evoluem com comprometimento da função respiratória o que, na maioria dos casos, leva à hospitalização. E o indivíduo de qualquer idade que apresente sintomas tais como febre alta, tosse, dispneia, mialgia, dor de garganta, saturação O_2 < 95% e desconforto respiratório, pode ir a óbito (Brasil, 2020).

A pandemia provocada pela Corona Virus Disease (COVID-19) teve origem no final de 2019, na capital da província de Hubei, Waham, na China, quando o vírus Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), uma variação de coronavírus humano, foi identificado. Isso tornou-se o novo desafio mundial, e impactou a vida da população nas mais diferentes perspectivas, mas de forma mais agressiva, nos campos de saúde e economia (Freitag et al, 2021).

Pacientes que apresentavam determinadas doenças crônicas respiratórias, cardíacas ou de natureza multifatorial tinham um prognóstico agravado quando eram apresentados ao vírus da COVID-19 (Yang et al., 2020). Neste sentindo é imprescindível determinar os principais grupos de risco especialmente tratando-se de uma pandemia, posto que essas informações facilitam a tomada de decisão dos profissionais da saúde (Alves et al.,2020).

As comorbidades são encontradas com frequência principalmente nos pacientes hospitalizados, destacando-se hipertensão arterial, diabetes e doença cardiovascular, sendo as duas últimas relacionadas com maior número de óbitos (Zhou et al., 2020). Das 265.099 mortes no Brasil por SRAG, referente ao período 30 de dezembro de 2019 a 05 de agosto de 2022, 194.192 foram devido à contaminação da COVID-19 (Brasil, 2022).

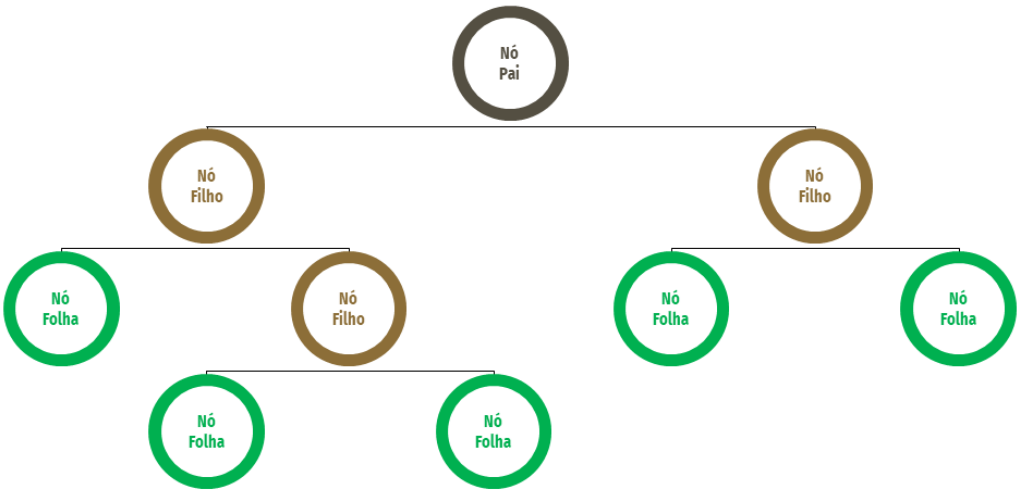
O presente trabalho tem como objetivo aplicar a modelagem preditiva via árvore de decisão (*decision tree*) para estimar a probabilidade: do indivíduo que tenha SRAG ser curado ou ir a óbito e do indivíduo que tenha SRAG ser curado ou ir a óbito devido à contaminação e não contaminação por COVID-19, analisando por fim os resultados encontrados (casos registrados no Brasil). Esses dados ajudarão os profissionais da saúde a entender como cada comorbidade se comportou/a.

2. Metodologia

2.1 Árvore de decisão

De acordo com Ragsdale (2001), a árvore de decisão é composta por nós (representados por círculos ou quadrados) interconectados por ramos (representados por linhas), essa intercalação é feita até que encontre uma folha, sendo esta a representação das classes. Um nó representa uma decisão, ramos emergindo do nó de decisão representam as diferentes alternativas para uma decisão particular. Essa representação pode ser vista na Figura 1 a seguir.

Figura 1 - Árvore de decisão.



Fonte: Autores (2022).

A estruturação de cada nó de uma árvore de decisão, vista na Figura 1, contém um teste sobre uma variável independente e os resultados desse teste formam os ramos das árvores. Os nós folhas/terminais representam valores de predição para as variáveis dependentes ou distribuições de probabilidade desses valores (Alves et al.,2020). O proposito básico da indução de uma árvore de decisão é produzir um modelo de predição acurado ou descobrir a estrutura preditiva do problema (Breiman et al.,1984). Conforme Wilkinson (2004) afirma: existem dois tipos de árvores de decisão, as árvores de regressão que tem sua variável dependente de valores numéricos e as árvores de classificação no qual as variáveis dependentes são categóricas.

Neste trabalho foi adotada a metodologia da poda da árvore de decisão por taxa de classificação, conforme James et al. (2013). Uma árvore de classificação é muito semelhante a uma árvore de regressão, exceto que é usada para prever uma resposta qualitativa em vez de quantitativa. O modelo de árvore de classificação pertence à categoria dos métodos de aprendizado de máquina por classificação supervisionada, como o modelo de regressão logístico (Albuquerque et al., 2020) ou redes neurais (Vogado, 2019), em que se deseja mapear uma variável dependente categórica Y com um conjunto de características ou variáveis independentes X.

Para uma árvore de regressão, a resposta prevista para uma observação é dada pela resposta média das observações de treinamento que pertencem ao mesmo nó terminal. Por outro lado, para uma árvore de classificação, estima-se que cada observação pertença à classe de treinamento mais comum entre as observações na região a qual à pertence (Grochtmann & Grimm, 1993).

Na interpretação dos resultados de uma árvore de classificação, geralmente se está interessado não apenas na previsão da classe correspondente a uma região de nó terminal específica, mas também nas proporções entre as observações de

treinamento que se enquadram nessa região (Loh, 2011).

A tarefa de cultivar uma árvore de classificação é bastante semelhante à tarefa de cultivar uma árvore de regressão. Assim como na configuração de regressão, utilizam-se recursos da divisão binária para aumentar uma árvore de classificação. No entanto, na configuração da árvore de classificação, a Soma de Quadrado do Resíduo (SQR) não pode ser usada como critério para fazer as divisões binárias.

Uma alternativa natural a SQR é a taxa de erro de classificação. Como se planeja atribuir uma observação em uma determinada região à classe de observações de treinamento mais comum nessa região, a taxa de erro de classificação é simplesmente a fração das observações de treinamento naquela região que não pertencem à classe mais comum:

$$E = 1 - \max_{k}(\hat{p}_{mk}), \tag{1}$$

em que $\hat{p}_{mk}$ representa a proporção de observações de treinamento na m-ésima região que pertencem à k-ésima classe. No entanto, verifica-se que o erro de classificação não é suficientemente sensível para o cultivo de árvores e, na prática, outras duas medidas são preferíveis (James et al., 2013).

Breiman et al. (1984) afirmam que o índice Gini é uma medida da variação total entre as classes K e é definido por

$$G = \sum_{k=1}^{K} \hat{p}_{mk}\,(1 - \hat{p}_{mk}), \tag{2}$$

Não é difícil ver que o índice Gini assume um pequeno valor se todos os estiverem próximos de zero ou um. Por esse motivo, o índice de Gini é referido como uma medida de pureza do nó, um pequeno valor indica que um nó contém predominantemente observações de uma única classe. Ou seja, um nó puro é um nó que tem predominância total de uma única classe. Moisen (2008) mostra que uma alternativa ao índice de Gini é a entropia cruzada, que é dada por

$$D = -\sum_{k=1}^{K} \hat{p}_{mk}\, log(\hat{p}_{mk}), \tag{3}$$

Como, $0 \leq \hat{p}_{mk} \leq 1$, segue-se que $0 \leq -\hat{p}_{mk} log(\hat{p}_{mk})$. Pode-se mostrar que a entropia cruzada assumirá um valor próximo de zero se os estiverem todos próximos de zero ou próximo a um. Portanto, como o índice de Gini, a entropia cruzada levará em um valor pequeno se o m-ésimo nó for puro. De fato, acontece que o índice Gini e a entropia cruzada são bastante semelhantes numericamente.

Ao construir uma árvore de classificação, o Índice de Gini ou a entropia normalmente são para avaliar a qualidade de uma determinada divisão, uma vez que essas duas abordagens são mais sensíveis à pureza do nó do que a taxa de erro de classificação (Rokach & Maimon, 2005). Qualquer uma das três abordagens pode ser usada quando se poda uma árvore, mas se a previsão da árvore final após a poda for o principal objetivo, a taxa de classificação é a preferível (James et al., 2013).

Algumas das vantagens identificadas em uma árvore de decisão são: representação gráfica intuitiva, útil em exploração de dados, menos limpeza de dados, manipulação variáveis numéricas e categóricas e é um método não paramétrico.

2.2 Os métodos de reamostragem

Os métodos de reamostragem são ferramentas indispensáveis na estatística moderna. Eles envolvem repetidamente amostras de um conjunto de treinamento e reajuste de um modelo. As abordagens de reamostragem eram computacionalmente custosas, porque elas envolvem o ajuste do mesmo método estatístico milhares de vezes usando diferentes subconjuntos dos dados de treinamento. Um dos métodos mais comuns de reamostragem é a Validação Cruzada (VC) (James et al.,2013).

De acordo com (Burman,1989) o método K-Fold de Validação Cruzada (VC) consistem em dividir o conjunto de observações em k grupos, ou dobras, de preferência com tamanhos iguais. Dos k grupos, um único grupo é retido como dados de validação para testar o modelo, e os k-1 grupos restantes são usados como dados de treinamento. Esse processo é repetido k vezes, com cada um dos k grupos usados exatamente uma vez como dados de validação, em cada um desses processos é calculado o Erro Quadrado Médio (EQM). Esse processo resulta em k estimativas do erro de teste, EQM 1, EQM 2, ..., EQM K. A estimativa da CV da dobra k é a por

$$CV_k = \frac{1}{k}\sum_{i=1}^{K} EQM, \tag{4}$$

no qual, quando a abordagem vai para o cenário de classificação a validação cruzada funciona exatamente como descrito anteriormente, exceto que em vez de usar o EQM para quantificar o erro de teste, usamos o número de observações classificadas incorretamente (James et al., 2013).

A validação cruzada não é capaz de aumentar a acurácia das estimativas, porém permite realizar uma média amostral (4) dos EQM, aumentando a certeza sobre o EQM.

2.3 Dados coletados para análise

Este estudo trata-se de uma pesquisa do tipo qualitativa longitudinal (Samperi et. al, 2013). Os dados foram coletados através do OpenDataSUS do Ministério da Saúde, no Brasil referente ao período de 2020 a agosto de 2022. Vale mencionar que o levantamento do material foi feito no dia 11 de agosto de 2022.

Vale destacar que o Ministério da Saúde (MS), por meio da Secretaria de Vigilância em Saúde (SVS), desenvolve a vigilância da SRAG no Brasil, desde a pandemia de Influenza A(H1N1). A partir disso, a vigilância de SRAG foi implantada na rede de vigilância, que anteriormente atuava exclusivamente com a vigilância sentinela de Síndrome Gripal (SG). Em 2020, a vigilância da COVID-19, foi incorporada na rede de vigilância da Influenza e outros vírus respiratórios (DataSUS, 2022).

A plataforma que coletamos os dados tem como finalidade disponibilizar o legado dos bancos de dados (BD) epidemiológicos de SRAG, desde o início da sua implantação (2009) até o ano corrente (2022). Atualmente, o sistema oficial para o registro dos casos e óbitos por SRAG é o Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVEP-Gripe) (DataSUS, 2022).

Ressaltamos que os dados coletados da vigilância de SRAG no Brasil disponibilizados, estão sujeitos a alterações decorrentes da investigação, ou mesmo correções de erros de digitação, pelas equipes de vigilância epidemiológica que desenvolvem o serviço nas três esferas de gestão. Esclarece-se que as bases de dados de SRAG disponibilizadas no portal passam por tratamento que envolve a anonimização, em cumprimento a Lei 13.709/2018 (DataSUS, 2022).

De acordo com Brasil (2022), os dados relacionados à SRAG devem ser lidos como mostra o Quadro 1 a seguir.

Quadro 1 - Dados analisados do banco de dados de Síndrome Respiratória Aguda Grave (incluindo dados da COVID-19).

COLUNAS	NOME DO CAMPO	DESCRIÇÃO	CATEGORIA
CARDIOPATI	Fatores de risco/ Doença Cardiovascular Crônica	Paciente possui Doença Cardiovascular Crônica?	1-Sim 2-Não
HEMATOLOGI	Fatores de risco/ Doença Hematológica Crônica	Paciente possui Doença Hematológica Crônica?	
SIND_DOWN	Fatores de risco/ Síndrome de Down	Paciente possui Síndrome de Down?	
HEPATICA	Fatores de risco/ Doença Hepática Crônica	Paciente possui Doença Hepática Crônica?	
ASMA	Fatores de risco/ Asma	Paciente possui Asma?	
DIABETES	Fatores de risco/ Diabetes mellitus	Paciente possui Diabetes *mellitus*?	
NEUROLOGIC	Fatores de risco/ Doença Neurológica Crônica	Paciente possui Doença Neurológica?	
PNEUMOPATI	Fatores de risco/ Outra Pneumatopatia Crônica	Paciente possui outra pneumopatia crônica?	
IMUNODEPRE	Fatores de risco/ Imunodeficiência ou Imunodepressão	Paciente possui Imunodeficiência ou Imunodepressão (diminuição da função do sistema imunológico)?	
RENAL	Fatores de risco/ Doença Renal Crônica	Paciente possui Doença Renal Crônica?	
OBESIDADE	Fatores de risco/ Obesidade	Paciente possui obesidade?	
CLASSI_FIN	Classificação final do caso	Diagnóstico final do caso. Se tiver resultados divergentes entre as metodologias laboratoriais, priorizar o resultado do RTPCR	1-SRAG por influenza 2-SRAG por outro vírus respiratório 3-SRAG por outro agente etiológico, qual: 4-SRAG não especificado 5-SRAG por COVID-19
EVOLUCAO	Evolução do caso	Evolução do caso	1-Cura 2-Óbito 3-Óbito por outras causas 9-Ignorado

Fonte: SIVEP-Gripe - Adaptado pelos autores (2022).

Os dados contidos no Quadro 1 foram encontrados no dicionário de dados do SIVEP-Gripe, disponibilizado pela SVS, vinculada ao MS. Tais dados foram utilizados na aplicação da modelagem preditiva via árvore de decisão (*decision tree*) para estimar a probabilidade: do indivíduo que tenha SRAG ser curado ou ir a óbito e do indivíduo que tenha SRAG ser curado ou ir a óbito devido à contaminação e não contaminação por COVID-19, analisando por fim os resultados encontrados (casos registrados no Brasil).

O banco de dados obtido no site do Ministério da Saúde, inicialmente apresentou 946.404 dados para o Brasil, contemplando informações do período 30 de dezembro de 2019 a 05 de agosto de 2022. No entanto, foi realizado um tratamento nesse banco visando deixar apenas os casos sinalizados pelo número 1 (sim) e 2 (não) nas colunas das comorbidades, e para isso foi feito a exclusão do número 9 (ignorado). Após o tratamento dos dados, o conjunto analisado considerou 926.387 dados para o Brasil, contemplando informações do período 30 de dezembro de 2019 a 05 de agosto de 2022. Chamaremos o conjunto analisado de conjunto A.

Os dados tratados do conjunto A passaram por um novo tratamento visando deixar somente os casos sinalizados pelo número 5 (SRAG por COVID-19), e para isso foi feito a exclusão do número 1 (SRAG por influenza), 2 (SRAG por outro vírus respiratório), 3 (SRAG por outro agente etiológico) e 4 (SRAG não especificado). Após o tratamento dos dados, o conjunto analisado considerou apenas 574.482 elementos para o Brasil, contemplando informações do período 22 de fevereiro de 2020 a 03 de agosto de 2022.

Os dados tratados do conjunto A passaram por outro tratamento visando deixar somente os casos sinalizados pelo número 1 (SRAG por influenza), 2 (SRAG por outro vírus respiratório), 3 (SRAG por outro agente etiológico) e 4 (SRAG não

especificado), e para isso foi feito a exclusão do número 5 (SRAG por COVID-19). Após o tratamento dos dados, o conjunto analisado considerou apenas 351.905 elementos para o Brasil, contemplando informações do período 30 dezembro de 2019 a 05 de agosto de 2022.

2.4 Softwares

Os Softwares utilizados para a construção e desenvolvimento desta pesquisa:

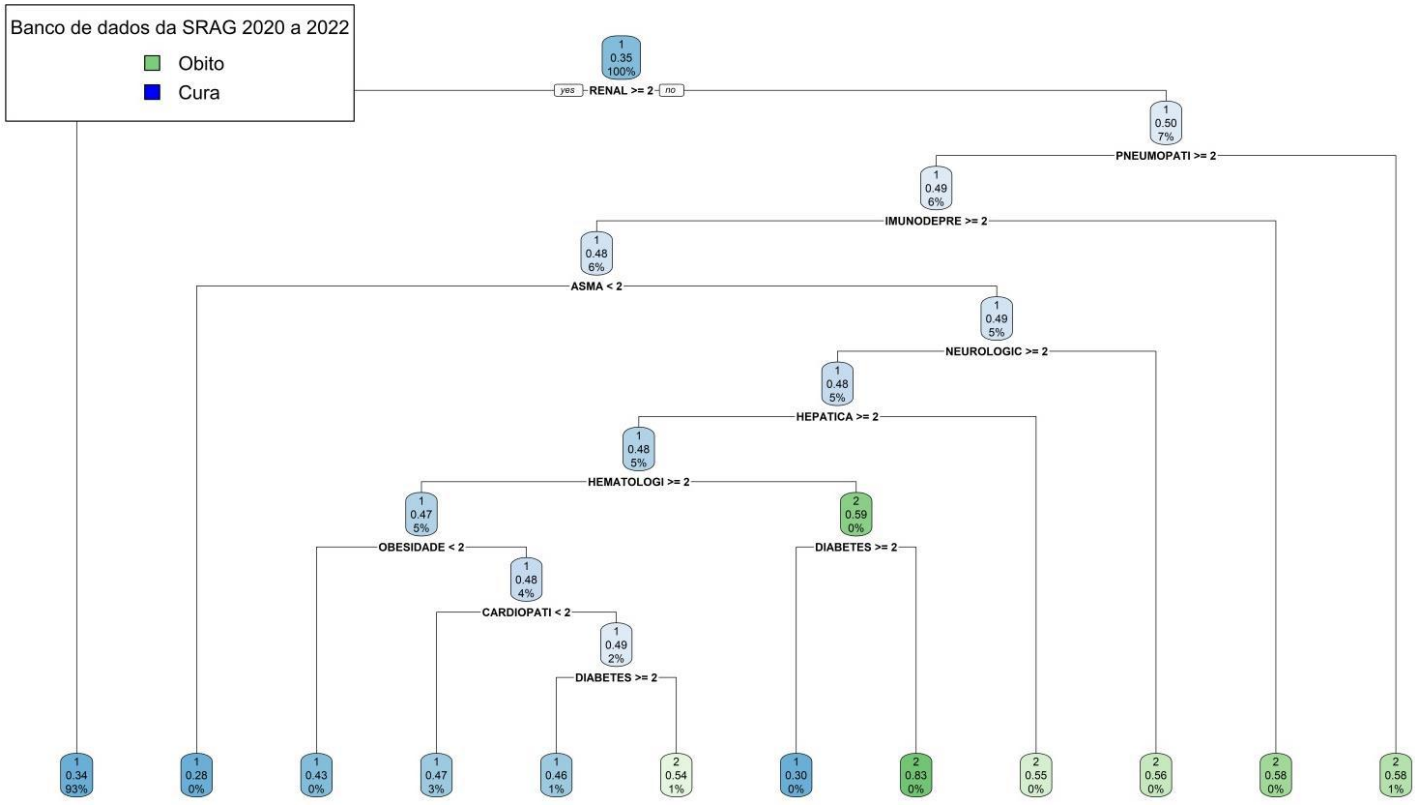
A organização dos dados foi feita no **Microsoft Office Excel** (2010), editor de planilhas produzido pela Microsoft para computadores que utilizam o sistema operacional Microsoft Windows, utilizados em computadores e dispositivos móveis para armazenamento de dados.

A análise estatística foi feita no **Software R** (2022), ambiente computacional com uma linguagem de programação que vem progressivamente se especializando em manipulação, análise e visualização gráfica de dados com ambiente disponível para diferentes sistemas operacionais: Unix/Linux, Mac e Windows.

3. Resultados e Discussão

Para analisar o comportamento dos dados do indivíduo que apresentou sintomas da SRAG. Foi-se utilizado o método de validação cruzada. O modelo final e de melhor precisão foi o modelo representado pela árvore de classificação da Figura 2.

Figura 2 - Probabilidade condicional das comorbidades: Renal Crônica, Pneumatopatia Crônica, Imunodeficiência ou Imunodepressão, Asma, Neurológica Crônica, Hepática Crônica, Hematológica Crônica, Obesidade, Cardiovascular Crônica e Diabetes mellitus – SRAG.



Fonte: Autores (2022).

Observa-se na árvore de decisão da Figura 2 as probabilidades dos indivíduos que tenham apresentado sintomas da

SRAG ser curado ou ir a óbito nos seguintes cenários:

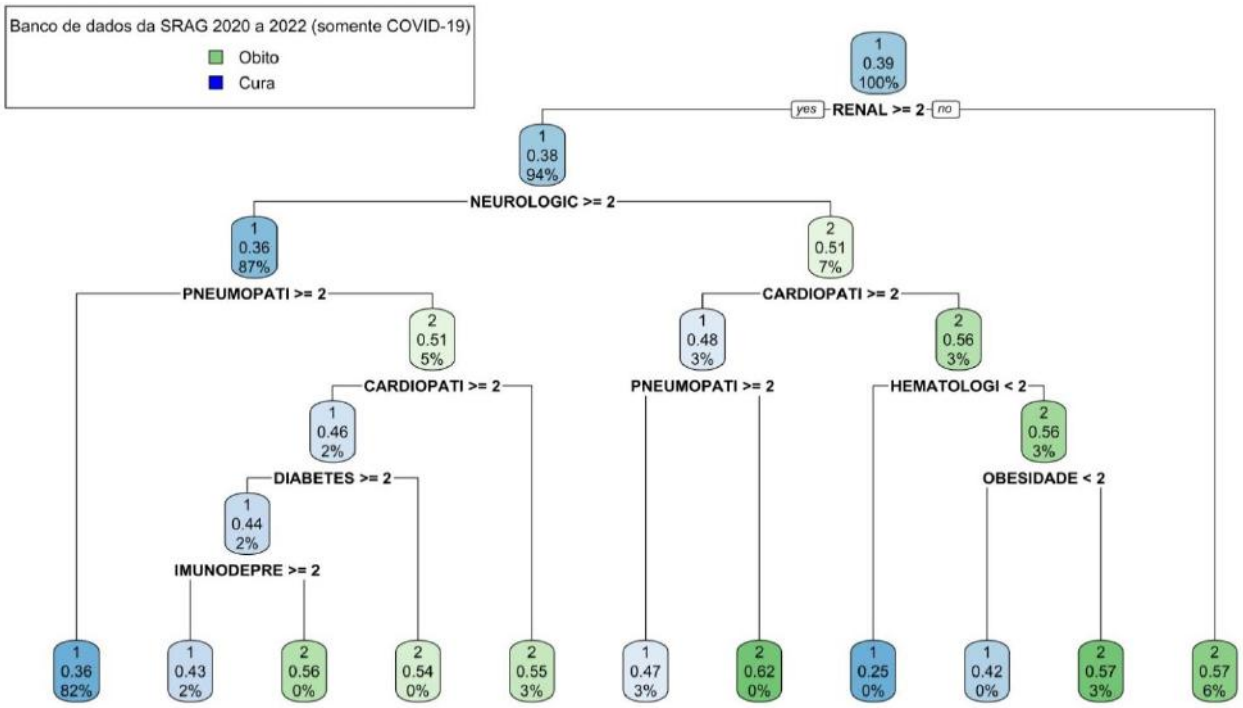
- Se o indivíduo apresentou sintomas da SRAG, então a sua chance de ser curado é de 65% e de ir a óbito é de 35%.
- Se o indivíduo que apresentou sintomas da SRAG tem doença renal crônica, então a sua chance de ser curado é de 50% e de ir a óbito é de 50%.
- Se o indivíduo que apresentou sintomas da SRAG tem doença renal crônica e pneumatopatia crônica, então a sua chance de ser curado é de 42% e de ir a óbito é de 58%.
- Se o indivíduo que apresentou sintomas da SRAG tem doença renal crônica e imunodeficiência ou imunodepressão, então a sua chance de ser curado é de 42% e de ir a óbito é de 58%.
- Se o indivíduo que apresentou sintomas da SRAG tem doença renal crônica e asma, então a sua chance de ser curado é de 72% e de ir a óbito é de 28%.
- Se o indivíduo que apresentou sintomas da SRAG tem doença renal crônica e doença neurológica crônica, então a sua chance de ser curado é de 44% e de ir a óbito é de 56%.
- Se o indivíduo que apresentou sintomas da SRAG tem doença renal crônica e doença hepática crônica, então a sua chance de ser curado é de 45% e de ir a óbito é de 55%.
- Se o indivíduo que apresentou sintomas da SRAG tem doença renal crônica e doença hematológica crônica, então a sua chance de ser curado é de 41% e de ir a óbito é de 59%.
- Se o indivíduo que apresentou sintomas da SRAG tem doença renal crônica, doença hematológica crônica e diabetes *mellitus*, então a sua chance de ser curado é de 17% e de ir a óbito é de 83%.
- Se o indivíduo que apresentou sintomas da SRAG tem doença renal crônica e obesidade, então a sua chance de ser curado é de 57% e de ir a óbito é de 43%.
- Se o indivíduo que apresentou sintomas da SRAG tem doença renal crônica e doença cardiovascular crônica, então a sua chance de ser curado é de 53% e de ir a óbito é de 47%.
- Se o indivíduo que apresentou sintomas da SRAG tem doença renal crônica e diabetes *mellitus*, então a sua chance de ser curado é de 46% e de ir a óbito é de 54%.

Os melhores cenários para os indivíduos que apresentaram sintomas da SRAG serem curados são hierarquicamente: ter doença renal crônica e asma, visto que a chance de ser curado é de 72%; não ter comorbidade, visto que a chance de ser curado é de 65%; ter doença renal crônica e obesidade, visto que a chance de ser curado é de 57%; ter doença renal crônica e doença cardiovascular crônica, visto que a chance de ser curado é de 53%; ter doença renal crônica, visto que a chance de ser curado é de 50%; ter doença renal crônica e diabetes *mellitus*, visto que a chance de ser curado é de 46%; ter doença renal crônica e doença hepática crônica, visto que a chance de ser curado é de 45%; ter doença renal crônica e doença neurológica crônica, visto que a chance de ser curado é de 44%; ter doença renal crônica e pneumatopatia crônica, visto que a chance de ser curado é de 42%; ter doença renal crônica e imunodeficiência ou imunodepressão, visto que a chance de ser curado é de 42%; ter doença renal crônica e doença hematológica crônica, visto que a chance de ser curado é de 41% e ter doença renal crônica, doença hematológica crônica e diabetes *mellitus*, visto que a chance de ser curado é de 17%.

Podemos concluir que é melhor que o indivíduo que apresentou sintomas da SRAG tenha doença renal crônica e asma do que não tenha comorbidade, visto que a chance de cura é maior 7%.

Após análise inicial, passamos a investigar o comportamento dos dados fixando os casos positivos para COVID-19. Sendo utilizado novamente o método de validação cruzada. O modelo final e de melhor precisão foi o modelo representado pela árvore de classificação da Figura 3.

Figura 3 - Probabilidade condicional das comorbidades: Renal Crônica, Neurológica Crônica, Cardiovascular Crônica, Pneumatopatia Crônica, Hematológica Crônica, Obesidade, Diabetes mellitus e Imunodeficiência ou Imunodepressão – SRAG (somente os dados da COVID-19).



Fonte: Autores (2022).

Observa-se na árvore de decisão da Figura 3 as probabilidades dos indivíduos que tenham apresentado sintomas da SRAG ser curado ou ir a óbito devido à **contaminação por COVID-19** dos seguintes cenários:

- Se o indivíduo que apresentou sintomas da SRAG (somente COVID-19), então a sua chance de ser curado é de 61% e de ir a óbito é de 39%.
- Se o indivíduo que apresentou sintomas da SRAG (somente COVID-19) tem doença renal crônica, então a sua chance de ser curado é de 43% e de ir a óbito é de 57%.
- Se o indivíduo que apresentou sintomas da SRAG (somente COVID-19) tem doença neurológica crônica, então a sua chance de ser curado é de 49% e de ir a óbito é de 51%.
- Se o indivíduo que apresentou sintomas da SRAG (somente COVID-19) tem doença neurológica crônica e doença cardiovascular crônica, então a sua chance de ser curado é de 44% e de ir a óbito é de 56%.
- Se o indivíduo que apresentou sintomas da SRAG (somente COVID-19) tem doença neurológica crônica, doença cardiovascular crônica e doença hematológica crônica, então a sua chance de ser curado é de 75% e de ir a óbito é de 25%.
- Se o indivíduo que apresentou sintomas da SRAG (somente COVID-19) tem doença neurológica crônica, doença cardiovascular crônica e obesidade, então a sua chance de ser curado é de 58% e de ir a óbito é de 42%.
- Se o indivíduo que apresentou sintomas da SRAG (somente COVID-19) tem doença neurológica crônica e pneumatopatia crônica, então a sua chance de ser curado é de 38% e de ir a óbito é de 62%.
- Se o indivíduo que apresentou sintomas da SRAG (somente COVID-19) tem pneumatopatia crônica, então a sua chance de ser curado é de 49% e de ir a óbito é de 51%.
- Se o indivíduo que apresentou sintomas da SRAG (somente COVID-19) tem pneumatopatia crônica e doença

cardiovascular crônica, então a sua chance de ser curado é de 45% e de ir a óbito é de 55%.

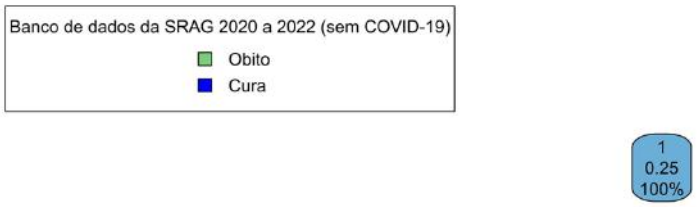
- Se o indivíduo que apresentou sintomas da SRAG (somente COVID-19) tem pneumatopatia crônica e diabetes *mellitus*, então a sua chance de ser curado é de 46% e de ir a óbito é de 54%.
- Se o indivíduo que apresentou sintomas da SRAG (somente COVID-19) tem pneumatopatia crônica e imunodeficiência ou imunodepressão, então a sua chance de ser curado é de 44% e de ir a óbito é de 56%.

Os melhores cenários para os indivíduos que apresentaram sintomas da SRAG devido à contaminação por COVID-19 serem curados são hierarquicamente: ter doença neurológica crônica, doença cardiovascular crônica e doença hematológica crônica, visto que a chance de ser curado é de 75%; nenhuma comorbidade, visto que a sua chance de ser curado é de 61%; ter doença neurológica crônica, doença cardiovascular crônica e obesidade, visto que a chance de ser curado é de 58%; ter doença neurológica crônica, visto que a chance de ser curado é de 49%; ter pneumatopatia crônica, visto que a chance de ser curado é de 49%; ter pneumatopatia crônica e diabetes *mellitus*, visto que a chance de ser curado é de 46%; ter pneumatopatia crônica e doença cardiovascular crônica, visto que a chance de ser curado é de 45%; ter doença neurológica crônica e doença cardiovascular crônica, visto que a chance de ser curado é de 44%; ter pneumatopatia crônica e imunodeficiência ou imunodepressão, visto que a chance de ser curado é de 44%; ter doença renal crônica, visto que a chance de ser curado é de 43% e ter doença neurológica crônica e pneumatopatia crônica, visto que a chance de ser curado é de 38%.

Podemos concluir que é melhor que o indivíduo que apresentou sintomas da SRAG devido à contaminação por COVID-19 tenha doença neurológica crônica, doença cardiovascular crônica e doença hematológica crônica do que não tenha comorbidade, visto que a chance de cura é maior 14%.

Após análise inicial, passamos a investigar agora o comportamento dos dados fixando os casos negativos para COVID-19. Sendo utilizado novamente o método de validação cruzada. O modelo final e de melhor precisão foi o modelo representado pela árvore de classificação da Figura 4.

Figura 4 - Probabilidade condicional das comorbidades – SRAG (sem incluir os dados da COVID-19).



Fonte: Autores (2022).

Observa-se na árvore de decisão da Figura 4 a probabilidade dos indivíduos que tenham apresentado sintomas da SRAG ser curado ou ir a óbito devido à **não contaminação por COVID-19** no seguinte cenário:

- Se o indivíduo que apresentou sintomas da SRAG (não contraiu o COVID-19), então a sua chance de ser curado é de 75% e de ir a óbito é de 25%.

Podemos concluir que o indivíduo que apresentou sintomas da SRAG, mas que não foi contaminado por COVID-19 tem a chance de cura de 75%.

4. Conclusão

Este trabalho buscou estimar a probabilidade: do indivíduo que tenha SRAG ser curado ou ir a óbito e do indivíduo que tenha SRAG ser curado ou ir a óbito devido à contaminação e não contaminação por COVID-19, analisando por fim os resultados encontrados (casos registrados no Brasil), utilizando como método a modelagem preditiva via árvore de decisão.

A partir da análise dos resultados encontrados do Brasil, referente ao período de 2020 a agosto de 2022, notou-se que o modelo proposto se ajusta bem as comorbidades, encontrando assim as porcentagens de sobrevivência, chegando as seguintes conclusões:

Nos casos que apresentaram sintomas da SRAG observou-se que o indivíduo tem a chance de 72% de ser curado se tiver doença renal crônica e asma e a chance de 65% de ser curado aquele que não tiver comorbidade. Logo, ter doença renal crônica e asma é melhor do que não ter comorbidade, visto que a sua chance de cura é maior 7%.

Nos casos que apresentaram sintomas da SRAG devido à contaminação por COVID-19 observou-se que o indivíduo tem a chance de 75% de ser curado se tiver doença neurológica crônica, doença cardiovascular crônica e doença hematológica crônica e a chance de 61% de ser curado aquele que não tiver comorbidade. Logo, ter doença neurológica crônica, doença cardiovascular crônica e doença hematológica crônica é melhor do que não ter comorbidade, visto que a sua chance de cura é maior 14%. Nos casos que apresentaram sintomas da SRAG devido à não contaminação por COVID-19 observou-se que o indivíduo tem a chance de 75% de ser curado.

Sugere-se para trabalhos futuros, utilizar a modelagem preditiva via árvore de decisão para uma análise em escalas regionais e estaduais do Brasil ou até mesmo de outro território nacional, analisando a semelhança entre as porcentagens de sobrevivência de modo a detectar possíveis relações entre eles, podendo modificar o espaço temporal da pesquisa, e investigar as correlações dos dados estatísticos com a área da saúde, buscando compreender o motivo (medicamentos, tratamentos, entre outros) para os resultados encontrados.

Referências

Albuquerque, M. A., Lucena, S. L. L., & Barros, K. N. N. O. (2020). Comparação de modelo clássico e Bayesiano para dados de óbitos perinatais no ISEA, Campina Grande - PB. *Research, Society and Developmen, 9*(8), e464985477. https://doi.org/10.33448/rsd-v9i8.5477

Alves, D. A. N. S., Nascimento, G. I. L. A., Castanha, E. R., Luna, J. E. L., Sobral, E. F. M., Brandão, W. A., Moreira, K. A., Mendes, J. S., Cunha Filho, M., Barros, D. M. & Falcão, R. E. A. (2020). Prevalência de comorbidades na Síndrome Respiratória Aguda Grave em pacientes com COVID-19 e outros agentes infecciosos. *Research, Society and Development, 9*(11), e70791110286. https://doi.org/10.33448/rsd-v9i11.10286

Brasil. (2022). Dicionário de dados. Ministério da Saúde. Secretária de Vigilância em Saúde. Sistema de Informação de Vigilância Epidemiológica da Gripe. https://s3.sa-east-1.amazonaws.com/ckan.saude.gov.br/SRAG/pdfs/dicionario_de_dados_srag_hosp_17_02_2022.pdf

Brasil. (2020). Protocolo de Manejo Clínico. Ministério da Saúde. Brasília, DF. https://www.saude.ms.gov.br/wp-content/uploads/2020/03/Protocolo-Manejo-Clinico_APS_versao04.pdf

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. CRC Press.

Burman, P. (1989). A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika, 76* (3), 503-514.

DataSUS. (2022). Ministério da Saúde. SRAG - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19. *OpenDataSUS*. https://opendatasus.saude.gov.br/dataset/srag-2021-e-2022

Freitag, V. L., Antonio, M. G. D., Loureiro, L. H., & Pereira, R. M. S. (2021). COVID 19 e a propagação de fake news sobre a contaminação pelo dióxido de carbono com o uso de máscaras faciais: Um estudo de reflexão. *Research, Society and Developmen, 10*(10), e104101018696. https://doi.org/10.33448/rsd-v10i10.18696

Grochtmann, M., & Grimm, K. (1993). Classification trees for partition testing. *Software Testing, Verification and Reliability, 3* (2), 63-82.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.

Loh, W. Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1* (1), 14-23.

Moisen, G. G. (2008). Classification and regression trees. *In: Jørgensen, Sven Erik; Fath, Brian D. (Editor-in-Chief). Encyclopedia of Ecology, volume 1. Oxford, UK: Elsevier*. 582-588., 582-588.

R Core Team. (2022). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. Vienna.

Ragsdale, C. T. (2001). *Spreadsheet modeling and decision analysis: a practical introduction to management science*. Cengage Learning.

Rokach, L., & Maimon, O. (2005). Top-down induction of decision trees classifiers-a survey. *IEEE Transactions on Systems, Man, and Cybernetics*, Part Applications and Reviews), 35(4), 476-487.

Samperi, R. H., Collado, C. F., & Lucio, M. del P. B. (2013). *Metodologia Científica*. AMGH Editora.

Vogado, L. H., Veras, R. M., Araujo, F. H., Silva, R. R., & Aires, K. R. (2019). Rede Neural Convolucional para o Diagnóstico de Leucemia. In Anais Principais do XIX Simpósio Brasileiro de Computação Aplicada à Saúde, 46-57.

Wilkinson, L. (2004). Classification and regression trees. *Systat*, 11, 35-56.

Yang, J., Zheng, Y., Gou, X., Pu, K., Chen, Z., Guo, Q., Ji, R., Wang, H., Wang, Y., & Zhou, Y. (2020). Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *International Journal of Infectious Diseases: IJID: official publication of the International Society for Infectious Diseases*, *94*, 91–95.

Zhou,F., Yu,T., Du,R., Fan,G., Liu,Y., Liu,Z., Xiang,J., Wang,Y., Song,B., Gu,X., Guan,L., Wei,Y., Li,H., Wu,X., Xu,J., Tu,S., Zhang,Y., Chen,H., & Cao,B. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. doi: 10.1016/S0140-6736(20)30566-3.